

Organización de la Información: Del catálogo a la Web Semántica

FILIBERTO FELIPE MARTÍNEZ ARELLANO
*Centro Universitario de Investigaciones Bibliotecológicas,
UNAM, México*

Para que los recursos de información sean conocidos, y por lo tanto utilizados, es necesario organizarlos adecuadamente. Si una biblioteca cuenta con una gran cantidad de recursos pero éstos no se encuentran organizados, su valor y utilidad pierden sentido. Contar con una gran cantidad de documentos o recursos de información sin organizar para proporcionar a los usuarios aquéllos que requieren, equivale prácticamente a no tener nada, pues será muy difícil encontrar los que el usuario necesita para satisfacer sus necesidades.

La necesidad de organizar los materiales bibliográficos, documentos o recursos de información que constituyen el acervo de las bibliotecas, ha estado presente a lo largo de su historia y evolución, adoptándose tres medios desde la aparición de éstas hasta nuestros días: la catalogación y los catálogos, los metadatos y la Web semántica. Cada una de estas alternativas posee sus propias características y su aparición ha sido motivada por la evolución de los medios para generar y distribuir información y las características de cada época.

CATALOGACIÓN Y CATÁLOGOS

La primera opción que fue utilizada para organizar los acervos o recursos de información de las bibliotecas fue la catalogación de los mismos y la inclusión de la información resultante dentro del catálogo. Es importante puntualizar que este instrumento fue desarrollado tomando en consideración las siguientes funciones básicas: localizar el material bibliográfico, un documento o cualquier otro recurso de información existente en la biblioteca o cualquier otra unidad de información para identificar si realmente es el que se necesita, así como valorar y tomar la decisión de su obtención.

La primera necesidad que tiene el usuario al asistir a una biblioteca y consultar el catálogo, o cualquier otro sistema de información, es conocer si la biblioteca posee determinado material o materiales que requiere, ya sea sobre determinado tema o autor o bien un material en particular. Es posible dar respuesta a lo anterior a través del establecimiento o creación de los denominados puntos de acceso, siendo los primordiales: autor, título y materia.

Una vez que los materiales han sido localizados, la siguiente necesidad de un usuario es identificar si esos materiales son los que realmente necesitan; el autor buscado, la edición requerida, el idioma en que se encuentra la información, el año en que fue impreso, etcétera. Todos estos elementos son incluidos en los registros del catálogo (anteriormente solo en tarjetas) a través de la descripción de las características principales de los materiales, basándose en estándares, normas o reglas de catalogación.

La identificación de los materiales le va a permitir al usuario valorar si éstos son de utilidad para sus necesidades y tomar la decisión de obtenerlos o no. La descripción de las características principales de los materiales o recursos de información también le va a permitir al usuario tomar la decisión sobre su obtención, por medio de los números de clasificación, o en nuestros días, a través de un URL.

Esta serie de funciones en que el catálogo se sustenta han estado presentes desde el catálogo de tarjetas, arregladas en ocasiones en muebles de madera preciosa, los que para los más jóvenes resultan raros, hasta los catálogos en computadora (OPACS), los que son más comunes en nuestros días.

La era en que actualmente vivimos tiene otra serie de características. Existen nuevas formas de generar, distribuir y acceder la información, el libro ha dejado de ser el único medio para llevar a cabo lo anterior y a éste se han sumado los medios audiovisuales, los recursos electrónicos y los multimedia, entre otros tipos de recursos de información. La era actual se caracteriza por una transformación de la comunicación impresa hacia una comunicación basada en medios electrónicos. Asimismo, Internet es utilizada comúnmente como una alternativa para producir, difundir, acceder y obtener información, a tal grado que se ha mencionado que las bibliotecas ya no van a ser necesarias, ni tampoco la catalogación y los catálogos.

Ciertamente, Internet y los documentos electrónicos representan un nuevo medio para generar, distribuir y obtener información. Sin embargo, al igual que los materiales o documentos impresos, para que los recursos de información electrónica sean conocidos, y por lo tanto utilizados, es necesario organizarlos adecuadamente. Lo anterior implica un reto para la organización de tales materiales o recursos de información. Dicho reto es el establecimiento y desarrollo de sistemas que permitan una efectiva organización de los recursos electrónicos para que éstos puedan ser localizados, identificados, valorados, accedidos (obtenidos) y utilizados por los miembros de las comunidades a las cuales va dirigida la información que contienen.

Es evidente que actualmente también es necesario que se continúen cumpliendo las funciones que se señalaron para el catálogo, pero ahora éstas tienen que ser trasladadas a un ambiente electrónico: localizar recursos electrónicos sobre determinado tema o autor; identificar si esos recursos son realmente los que el usuario necesita; valorar su utilidad para tomar la decisión sobre su obtención. En este último caso, la tecnología actual nos brinda un gran apoyo, ya que ahora podemos acceder más fácilmente un recurso de información, no importando el lugar donde éste se encuentre, Japón, China, Australia o Argentina, por mencionar algunos sitios, lo que anteriormente era más difícil.

METADATOS

Actualmente, además de la catalogación y los catálogos, existen otras alternativas para organizar y localizar los materiales bibliográficos, documentos o cualquier otro tipo de recursos de información, siendo una de ellas los metadatos.

Este término da la impresión de que se trata de algo nuevo y de que el catálogo ha sido superado, sin embargo los metadatos se basan en los mismos principios y funciones en que se han sustentado la catalogación y el catálogo, aunque también presentan algunas diferencias, producto del desarrollo tecnológico y de las necesidades de los usuarios de nuestros días.

Como ha sido mencionado anteriormente, los medios para generar, distribuir y obtener información se han ampliado a los recursos electrónicos, los cuales han traído consigo una serie de problemas para su organización, siendo algunos de ellos: su crecimiento exponencial, su movilidad frecuente e inestabilidad, la diversificación de sus tipos y formatos (existencia de recursos textuales y otros con imágenes como fotografías, mapas y planos, así como películas y video digitales), la existencia de necesidades de información diferentes (sector académico, educación en línea, comercio, gobierno, industria, etcétera). Adicionalmente, para la búsqueda y localización de este tipo de recursos de información se han desarrollado motores de buscadores y metabuscadores, los cuales se han basado en la utilización de metadatos.

Si comparamos la catalogación y los catálogos con las nuevas formas de la organización de la información, es factible notar que existen una serie de similitudes pero también diferencias. En la catalogación tradicional, los objetos a ser descritos eran tangibles: libros y documentos. En la organización de los recursos electrónicos, los objetos de ésta son intangibles: documentos conceptualizados y denominados ahora como objetos de información digital. Los procesos en la catalogación tradicional estaban centrados en la catalogación descriptiva y la clasificación de libros y documentos tangibles, utilizando estándares. En la catalogación de los recursos electrónicos, estos procesos están enfocados hacia la creación de metadatos descriptivos y metadatos orientados hacia su contenido, los cuales también están basados

en estándares. Por otro lado, los productos de la catalogación son los registros bibliográficos y los catálogos, anteriormente los de tarjetas y en nuestros días los catálogos en línea, OPACS, de acuerdo a los términos en inglés, o los denominados WebPAC, catálogos en la Web. En la organización de los recursos electrónicos, los productos son registros y archivos de metadatos, así como sistemas de recuperación basados en metadatos (Méndez Rodríguez).

La creación de los registros en la catalogación tradicional ha estado basada en estándares tales como las ISBD (*International Standard Bibliographic Description*), las RCA2 (Reglas de Catalogación Angloamericanas, segunda edición) y el formato MARC. Dadas las similitudes existentes entre la catalogación tradicional y la catalogación de los recursos electrónicos se han hecho adaptaciones a dichos estándares, tales como las ISBD (ER) (*International Standard Bibliographic Description for Electronic Resources*), la actualización del capítulo 9 para recursos electrónicos de las RCA2 y actualmente la creación de las RDA (Resource Description and Access). No obstante lo anterior, las diferencias existentes también han motivado el surgimiento de los metadatos.

Los metadatos han sido definidos de diversas formas, desde la definiciones mas simples: datos acerca de datos o información acerca de información; o conjunto de elementos que pueden ser usados para describir y representar objetos de información, las cuales también podrían ser aplicables a la catalogación descriptiva, hasta otras en donde se toman en cuenta la naturaleza y las características de los recursos de información electrónica.

Los metadatos también han sido definidos como datos asociados con objetos de información, los cuales permiten a sus usuarios potenciales tener un conocimiento previo de su existencia y/o sus características (Dempsey y Heery). Datos que describen los atributos de un recurso de información y apoyan su localización, identificación, evaluación y selección, entre otras funciones, las cuales pueden ser llevadas a cabo por un usuario final o por sus agentes (otras personas o sistemas automatizados) (Dempsey y Heery). Es conveniente señalar que en la definición anterior existen similitudes con las funciones básicas de la catalogación y el catálogo, pero también diferencias en el sentido de

que éstas pueden ser llevadas a cabo por un ser humano o por agentes y sistemas automatizados, aquí es donde podemos empezar a notar una separación entre la catalogación y los metadatos.

Otra definición de los metadatos señala que son datos con sentido propio que proporcionan información o documentación acerca de otros datos manejados dentro de una aplicación o ambiente... Los metadatos pueden incluir información descriptiva acerca del contexto, calidad y condiciones o características de los datos (Taylor). En esta última definición se puede observar que los metadatos pueden incluir datos acerca del contexto de la información o de un documento, los que anteriormente estaban relegados, centrándose la catalogación exclusivamente en la descripción de las características de un documento.

La conceptualización anterior de los metadatos ha generado una tipología de éstos en tres grandes grupos: los metadatos descriptivos, los metadatos estructurales y los metadatos administrativos.

El primer grupo, los metadatos descriptivos, tiene como objetivo primordial la descripción de los recursos de información para su localización e identificación. Estos se centran en la descripción de los atributos bibliográficos de un recurso de información (título, autor/ creador, idioma, palabras claves), sus atributos físicos (medios, condición de las dimensiones) y datos de identificadores únicos (PURL, Handle). En un nivel local, permiten la búsqueda y la recuperación de recursos de información específicos (por ejemplo: búsqueda en una colección de imágenes para encontrar pinturas con ilustraciones de animales). A nivel Web, permiten a los usuarios descubrir un grupo recursos de información (por ejemplo: búsqueda en la Web para encontrar colecciones digitalizadas sobre poesía). Estos aspectos también se encuentran presentes en los catálogos, pero existen otros que son cubiertos de mejor forma por los metadatos estructurales y administrativos.

Los metadatos estructurales facilitan la navegación y presentación de los recursos electrónicos. Asimismo, proporcionan información sobre la estructura interna de este tipo de recursos, incluyendo datos sobre una página, sección, capítulo, numeración, índices y tabla de contenido. Un libro impreso es un objeto único, pero en un libro digital, cada uno de sus capítulos conforman otra unidad independiente y al navegar de un capítulo a otro estamos manejando varios objetos

dentro de un mismo objeto. Este tipo de metadatos describen la relación entre los materiales (por ejemplo, la fotografía B fue incluida en el manuscrito A); asimismo, unen los archivos y los textos relacionados (por ejemplo, el Archivo A es el formato JPEG de la imagen de archivo del Archivo B).

La catalogación tradicional no puede resolver estos problemas; sin embargo, los metadatos estructurales permiten establecer la relación existente entre diferentes objetos, por ejemplo, entre las fotografías incluidas las diferentes partes de un recurso de información digital como puede ser un libro o un documento digital. Este tipo de metadatos utilizan etiquetas para las diferentes partes de la estructura de los documentos, para la página del título, la tabla de contenido, capítulos, partes, fe de erratas o el índice. Estos son manejados a través de distintos estándares de codificación como SGML, XLM, *Encoded Archival Description*, EAD (Descripción de Archivos Codificada) MOA2, *Electronic Binding (Ebind)*.

Por otro lado, los metadatos administrativos facilitan la gestión y procesamiento de las colecciones de recursos digitales, tanto a corto como a largo plazo, incluyen datos técnicos sobre la creación y el control de calidad de los recursos de información. También incorporan datos para el manejo de los derechos de autor y requisitos para el control de su acceso y utilización, además de información sobre acciones de preservación. Los metadatos administrativos comprenden datos técnicos tales como tipo y modelo de escáner, resolución, profundidad de bit, espacio de color, formato de archivo, compresión, fuente de luz, propietario, fecha del registro de derecho de autor, limitaciones en cuanto al copiado y distribución, información sobre licencia, actividades de preservación.

Tomando en consideración todo lo anteriormente expuesto es factible e importante mencionar que los metadatos son utilizados para la descripción y la localización de documentos o recursos de información. Pero además éstos son utilizados por robots y agentes de software, el comercio electrónico, firmas digitales. Adicionalmente, permiten salvaguardar los derechos de propiedad intelectual, así como la valoración, evaluación y clasificación de los contenidos de los recursos de información electrónica.

WEB SEMÁNTICA

No obstante que los metadatos han contribuido a la organización de la información electrónica, continúan existiendo una serie de problemas para ésta, debido, entre otras, a las siguientes situaciones: existencia de una sobreoferta de información, gran heterogeneidad de las fuentes de información y la dificultad en la interoperabilidad de los sistemas. Lo anterior provoca búsquedas de información con resultados ineficaces e imprecisos (Hurtado Bustamante). Asimismo, el método de marcado de la información (html) combina elementos de contenido con otros de presentación. Para un ser humano no hay ningún problema en interpretar el título de un documento a partir, por ejemplo, de su preeminencia, su formato y su lugar en la página, pero si el autor ha marcado el título con un elemento de formato () en lugar de uno semántico (<h1>), para la computadora resultará imposible identificar el título (Pedraza-Jiménez, Codina y Rovira).

Tomando en consideración la forma en que se recupera la información, actualmente se habla de dos tipos de Web, la sintáctica y la semántica. En la Web sintáctica, las computadoras realizan la presentación visual de la información (tarea fácil) y las personas navegan e interpretan su contenido (tarea difícil). ¿Sería posible que las computadoras hiciesen algo más? La respuesta es sí, a través de la Web semántica, la cual ha sido definida como una Web extendida, dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas específicas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida. En la Web actual existen datos que sólo son accesibles para humanos; en la Web semántica tenemos información en vez de datos, la cual es accesible a las personas, pero también a las máquinas a través de la lectura de los datos, y lo más importante, por medio de su interpretación (Rodríguez).

La Web semántica permite a los usuarios delegar tareas en el software. Gracias a la semántica, el software es capaz de procesar su contenido, razonar con éste, combinarlo y realizar deducciones lógicas para resolver problemas cotidianos de recuperación de la información en forma automática, con resultados más precisos, como en el ejemplo mostrado a continuación:

Figura 1: resultados obtenidos con un buscador normal



Fuente: Hurtado Bustamante, "Introducción a la Web semántica: realidades y perspectivas"

En este ejemplo, la necesidad de información de la que se partió es “vuelos a Praga para mañana por la mañana”. Como puede ser observado, con un buscador actual se recupera información que es irrelevante y esto es precisamente lo que en muchas ocasiones acontece al utilizar diferentes motores de búsqueda. Por otro lado, con un motor semántico los resultados son más precisos, obteniendo la información que se necesita.

La Web semántica implica el uso y desarrollo de lenguajes que faciliten la introducción de contenidos legibles por las máquinas. La Web semántica se basa en dos conceptos fundamentales: *Descripción* y *manipulación* del significado de la información. La descripción requiere de semántica, metadatos y ontologías. La manipulación, lógica y motores de inferencia (Hurtado Bustamante).

Tradicionalmente, la semántica ha sido relacionada con el estudio del significado de los términos lingüísticos, pero en este contexto significa la dotación de significado interpretable por parte de las máquinas. Los lenguajes de marcado como el XML, no añaden semántica a

un documento, siendo necesarios los metadatos y las ontologías. Los metadatos son los datos que describen los recursos de la Web. Las ontologías implican un conjunto de términos, además de las relaciones entre los términos que describen un dominio de aplicación concreto. Su objetivo es la creación de un diccionario de términos compartidos y comprensibles para diferentes aplicaciones y/o comunidades que posibiliten la interoperabilidad, así como compartir información y conocimiento entre ellas (Rodríguez).

En otras palabras, las ontologías son una forma de representación del conocimiento y de la información. Son la base para la arquitectura de la Web semántica. Un vocabulario aceptado por una comunidad amplia. Pueden inferir conocimiento implícito a partir de conocimiento representado (Hurtado Bustamente).

Por otro lado, la lógica es un razonamiento, método o sentido común. Un motor de inferencia extrae conclusiones a partir de un conjunto de reglas o un conjunto de premisas o axiomas. Existen motores de inferencia basados en diferentes órdenes de lógicas, métodos de resolución de problemas y algoritmos especializados que infieren dentro de sistemas expertos soluciones *ad-hoc* (Rodríguez).

CONSIDERACIONES FINALES

Indudablemente, los métodos y estándares para organizar la información han cambiado, adaptándose a las características de los diferentes tipos de materiales que han surgido, así como han incorporado el uso de las tecnologías. Sin embargo, los principios sustanciales en los que descansa la organización y recuperación de la información continúan siendo los mismos. Localizar la información, identificar si es la información que necesitamos para tomar la decisión de obtenerla. En estos principios descansan desde el desarrollo de los catálogos hasta el de la Web semántica.

No obstante lo anterior, actualmente la organización de la información ha traído consigo una serie de retos para los que nos dedicamos a esta área de nuestra disciplina. El primero de ellos es adaptarse al cambio, así como la necesidad de adquirir nuevas competencias técnicas y

tecnológicas. Pero además de lo anterior, es necesaria una introspección de nuestras competencias para organizar la información. Los catalogadores tenemos toda una serie de potencialidades, un cúmulo de conocimientos y experiencias. Los principios de la organización de la información siempre han estado presentes, pero los cambios producidos por la tecnología han motivado que éstos se apliquen en otro tipo de materiales o documentos, en instrumentos de recuperación apoyados en la tecnología. Si los organizadores de la información logramos afrontar estos retos, tenemos un futuro muy amplio y promisorio.

REFERENCIAS Y OBRAS CONSULTADAS

Dempsey, L. y Heery, R. "A review of metadata: a survey of current resource description formats". En: <http://www.ukoln.ac.uk/metadata/desire/overview/> (Consultado, 24 de noviembre de 2008).

Hurtado Bustamante, D. P. "Introducción a la Web semántica: realidades y perspectivas". En: http://dintev.univalle.edu.co/pierrelevy/documentos/otros/IntroduccionALaWebSemantica_DianaPaolaHurtado.pdf (Consultado, 24 de noviembre de 2008).

"Guía Breve de Web Semántica". En World Wide Web Consortium, Oficina Española. <http://www.w3c.es/divulgacion/guiasbreves/WebSemantica> (Consultado, 24 de noviembre de 2008).

Howarth, L. C. "Facilitar Metadatos: crear registros básicos para la búsqueda de recursos". En: World Library and Information Congress: 70th IFLA General Conference and Council, 22-27 August 2004, <http://www.ifla.org/IV/ifla70/prog04.htm> (Consultado, 24 de noviembre de 2008).

Lozano Tello, A. "Ontologías en la Web Semántica". En: I Jornadas de Ingeniería Web' 01, [http://www.informandote.com/jornadasIng WEB/articulos/jiw02.pdf](http://www.informandote.com/jornadasIng_WEB/articulos/jiw02.pdf) (Consultado, 24 de noviembre de 2008).

Méndez Rodríguez, E. M. "Catalogación / organización de documentos digitales: estado de la cuestión, tendencias y perspectivas desde España". En: E-Prints Library and Information Science, http://eprints.rclis.org/8088/1/Eva-Mendez_Anuari03.pdf (Consultado, 24 de noviembre de 2008).

Pedraza-Jiménez, R., Codina, L. Y Rovira, C. (2007). "Web semántica y ontologías en el procesamiento de la información documental". En: *El profesional de la información* 16(6). Barcelona: Swets & Zeitlinger. p. 569 - 578.

Rodriguez, H. "Web Semántica". En: [http://ecibjaveriana.wordpress.com/category/ Web-semantica/](http://ecibjaveriana.wordpress.com/category/Web-semantica/) (Consultado, 24 de noviembre de 2008).

Taylor, A. *Organization of information*. Englewood, Colo.: Libraries Unlimited, 2004.

Torres Pombert, A. "¿Catalogación en el entorno digital?: una breve aproximación a los metadatos". En: *Acimed* 2006, 14(5). http://bvs.sld.cu/revistas/aci/vol14_5_06/aci09506.htm (Consultado, 24 de noviembre de 2008).